Title: **Application of Tree-based Machine Learning Models for Energy Performance Certificates Predictions of Dwellings in The Netherlands**

Diego Tejada Arango, Isaï Magan, Arjan Zwamborn

Dutch Organisation for Applied Scientific Research (TNO)

The European Union Energy Performance of Buildings Directive (EU/2024/1275) requires an increase in energy performance by 16% by 2030 compared to 2020. This makes a standardised energy labelling system crucial to determining the energy efficiency of dwellings across Europe. Furthermore, it allows for identifying the need for energy efficiency improvements in the residential building sector in EU member states. At least 55% of the decrease in the average primary energy use must be achieved by renovating the worst-performing buildings. It is, therefore, crucial to understand how such renovations impact the envisioned standardised energy labelling system.

An energy labelling system has already been implemented in the Netherlands. However, approximately 42% of dwellings in the Netherlands still need an energy performance certificate (EPC) assignment. Labelling all these dwellings can be a cumbersome process. In the Dutch system, labelling is only required when a building is sold or rented. In this research, we analyse whether the EPC can be predicted based on the characteristics of already labelled dwellings, such as their shape, year of construction, insulation, and more.

Furthermore, we investigate the contribution of different renovation possibilities to the label assignment of the dwellings. This allows us to determine the most influential energy efficiency improvements and how adjustments in existing dwellings could lead to altered EPC assignments.

To tackle the missing EPC problem, we propose a machine-learning model. The model learns from a part of the Dutch energy label database to develop prediction models, which are then tested on another part of the database. Specifically, we apply tree-based machine learning models such as random forest and boosted trees models.

For the interpretability of the model results, we use Shapley values. These values allow us to rank and quantify feature importance for the EPC assignment. Specifically, this concerns all characteristics considered when training the model. We distinguish between two types of characteristics: actionable and non-actionable characteristics, where actionable characteristics can be altered through renovation measures.

This study shows that tree-based machine learning models like random forest and XG-Boost are powerful tools for recognizing patterns in the existing Dutch energy performance certificate databases. The Shapley values show different contributions of each building characteristic for the EPC assignment. For the actionable characteristics, we find that renovation measures such as improving the insulation quality of the roof and façades and switching to a higher efficiency boiler system improve the EPC assignment the most for dwellings with a poor EPC assignment.

The average proportion of all classifications that were correctly identified in the test set (i.e., the accuracy metric) is around 80% across all the EPCs, with higher accuracy for classifying labels A, C, and G. Classifying intermediate labels has proven more difficult since the criteria for determination are more diffuse among contiguous intermediate energy labels. A confusion matrix illustrating the accuracy of each possible EPC assignment is presented in Figure 1.

Figure 1. Confusion matrix of the percentage of predicted EPC assignments against actual EPC assignments

In addition, we perform an interpretability analysis that shows the most important features that determine each EPC. For instance, the compactness of the dwelling, the insulation level of the outer walls and windows and the year of construction are common factors in determining the labels. Two interpretability plots for dwellings with the worst EPC assignment can be found in Figure 2.

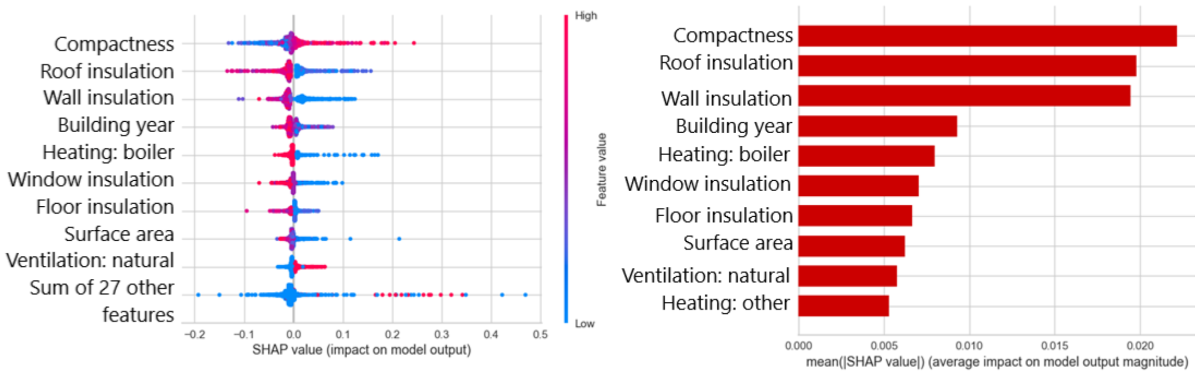

Figure 2. Interpretability plots depicting the distribution of SHAP values (left) and average SHAP value (right) for the most impactful building characteristics of dwellings with the worst performing EPC assignment

The interpretability analysis has shown that among the actionable characteristics, the insulation quality of the roof and facades and the type of space heating are, on average, the most impactful for the worst-performing EPC assignments.

Our findings open the door to using machine learning models to predict and explain the assignment of energy labels to dwellings in the Netherlands that still need to be assigned an EPC. This would give policymakers a practical tool to assess the progress made in achieving the targets outlined by the EU Commission. Additionally, it could provide a robust tool for measuring what policy measures could be most effective at improving the EPC assignments of the worst-performing dwellings, even if EPC assignment criteria change in the future. Finally, the proposal is sufficiently generalised that it can be applied to other EU member states as well, effectively standardising the energy performance of dwellings and how this could be improved. Among others, Sesana et al. (2024), Sayfikar & Jenkins (2024) and Aitziber et al. (2025) have concluded that there is a need for such standardisation.

The work presented in this abstract is funded by the Netherlands Enterprise Agency (Rijksdienst voor Ondernemend Nederland - RVO).

References

Aitziber, M., Cruz, N., Leandro, F., & Oxana, S. (2025). The future of EPCs: data, policy, and public perception in the race for energy efficiency. *Urban Governance*. https://doi.org/10.1016/j.ugj.2025.02.006

Sayfikar, M., & Jenkins, D. P. (2024). Energy performance certificate calculation methodologies across Europe and accommodating new performance indicators. *Building Services Engineering Research and Technology*. https://doi.org/10.1177/01436244241282076

Sesana, M. M., Salvalai, G., Della Valle, N., Melica, G., & Bertoldi, P. (2024). Towards harmonising energy performance certificate indicators in Europe. *Journal of Building Engineering*, *95*, 110323. https://doi.org/10.1016/j.jobe.2024.110323